Student Name: _____

Math 155 — Epidemiology

Review Questions

Problem 1

In 1970, the crude mortality rate (all causes) for Guyana (a developing country) was 6.8 per 1000 and for the United States it was 9.4 per 1000.

(a) Can the lower crude mortality rate in Guyana be explained by the fact that the United States has a larger population? Explain your answer. Answer comment:No.

The mortality rate adjusts for the size of the population.

(b) Give a probable explanation for the lower crude mortality rate in Guyana. Answer comment:Assuming that the lower rate in Guyana is not due to healthy living conditions or good medical care (which would seem unlikely in a developing country), a reasonable explanation is that the population of Guyana tends to be younger than that of the United States.

Problem 2

Crude and age-adjusted mortality rates (per 100,000 people) from "arteriosclerotic and degenerative heart diseases" are shown for Chile and the United States for 1967.

Country	Crude Rates	Age-Adjusted Rates
Chile	67.4	58.2
United States	316.3	131.4
Ratio: US/Chile	4.7	2.3

Which of the two rates is preferable for comparing the mortality rate from heart disease in the two countries? Why? Why do the ratios of the crude and age-adjusted rates for the two countries differ?

Answer comment:In general, age-adjusted rates are much preferred to crude rates. The crude rate does not take into account the age distribution of the population, which can vary markedly between countries (and between eras within a country). The crude and age-adjusted rates differ when the age distribution differs.

Problem 3

Assume that the prevalence of coronary heart disease decreases after age 70, but its incidence continues to increase with age. What is the most probable explanation for the divergence of these rates?

Answer comment:Since prevalence is incidence times duration, one way to reduce the prevalence is to reduce the duration. So, the decrease in prevalence in the presence of an increase in incidence signals a decrease in duration. Older people survive less time after getting CHD than younger people.

Problem 4

Two thousand women aged 55 years were given a health check and 100 were found to have high blood pressure. Ten years later all 2000 women attended a second check and another 300 women had developed high blood pressure.

(a) What was the prevalence of high blood pressure in the women (i) at age 55 and (ii) at age 65? Answer comment:Prevalence at age 55: 100/2000

Prevalence at age 65: 400/2000 (assuming the original 100 still had high blood pressure.

(b) How many women were "at risk" of developing high blood pressure at the start of the 10-year period? Answer comment:Since 100 people already had it, only the remaining 1900 were at risk for developing high blood pressure.

(c) What was the incidence of high blood pressure in these women? Is this a measure of cumulative incidence or an incidence rate? Answer comment:300 out of 1900 at risk got high blood pressure over the 10 year period. That's a cumulative incidence of $\frac{300}{1900}/10$, or 15.8/1000 per year.

It's a cumulative incidence rather than an incidence rate because we have not taken into account when each person got high blood pressure. For an incidence rate, whenever someone gets high blood pressure, we stop accumulating their years at risk.

Assume that, on average, each of the 300 women who developed high blood pressure did so half-way through the ten year follow-up period.

(d) Calculate the total number of person-years at risk (of developing high blood pressure) during the 10 years.

Answer comment:The assumption tells us that the 300 women who got high blood pressure were (on average) at risk for only 5 years of the 10 year period. That's 1500 person years-at risk. In addition, the $1900 - 300 = 1600$ women who were at risk but never got high blood pressure contribute 16000 person-years at risk. Total: 17500 person-years at risk

(e) What was the incidence rate of high blood pressure in these women? Answer comment:300 people per 17500 person-years at risk gives an incidence rate of 17.1/1000. This is slightly higher than the cumulative incidence.

Problem 5

Four rats in one cage were found dead at a university animal colony. In the adjacent cage, one rat convulsed and died, two rats became ill but survived, and one rat was ill. The veterinarian declared that an epizootic was present (epizootic — epidemic in animals).

Which type of study was this:

[**a**] Uncontrolled observation
[B] Cross-sectional or prevalence study
[C] Experiment
[D] Cohort Study
[E] Case-control study

Problem 6

The study hypothesis is that alcoholics have an increased incidence of fatal automobile accidents. Design a case-control study to test this hypothesis using the following headings:

(a) Diagnosis of cases — difficult or not? Where would you find cases?

(b) Name a suitable population from which to choose controls.

(c) List matching characteristics for controls.

(d) What essential characteristic must you determine for each study member, and what difficulties might be encountered in determining this characteristic?

Problem 7

The likelihood of a pathogen or agent being transmitted from one infected person to another susceptible person is referred to as which of the following?

invasiveness variability virulence ***communicability***

Problem 8

Which of the following terms refers to a persistent, expected level of disease in a defined population?

epidemic outbreak ***endemic*** pandemic

Problem 9

Serial surveys involve which of the following?

[A] Ecologic data
[B] Cross-sectional data
[C] Case-control data
[**d**] Cohort data

Problem 10

Match the following as a **descriptive** or an **analytic** epidemiological study.

- Who ***descriptive*** analytic
- Why descriptive ***analytic***
- How descriptive ***analytic***
- What ***descriptive*** analytic
- When ***descriptive*** analytic
- Where ***descriptive*** analytic

Problem 11

The outbreak of an illness among football players and coaches is hypothesized to be associated with a person's location — the north versus the south wing — in an athletic training center. Each player and coach was asked where he was practicing, resulting in the following 2×2 table.

	Illness	
Location	Yes	No
South	25	13
North	5	15

Use an appropriate measure to calculate the association between location and illness.

1.2 ***2.6*** 3.0 4.9 5.0

Problem 12

Suppose you wanted to monitor the trend in age-adjusted incidence rates of breast cancer over the calendar years 1980 through 2000. Which of the following standard populations is the most appropriate for adjusting the rates throughout this period?

A 1980
B 1990
C 2000
d It does not matter.

Problem 13

In Los Angeles, public health authorities identified 100 people with disease X. These cases came from 40 affected households with 230 people. Assuming that each household had only one primary case and that each primary case attended a local daycare center, what is the secondary attack rate?

a 32%
B 17%
C 43%
D 67%

Answer comment:There were 40 primary cases, leaving $230-40 = 190$ people susceptible. Assuming that, of the 100 cases, all but the 40 primary cases are secondary cases (the problem statement gives no choice but to make this assumption), then the secondary attach rate is $60/190 = 32\%$.

Problem 14

Answer these true-or-false questions.

1. Cancer screening is a form of primary prevention. True ***False***
2. The ecologic study design is a useful approach for measuring the association between exposure and disease at the individual level. True ***False***
3. Incidence is a better measure of disease risk than prevalence. True ***False***
4. If the disease is common, the odds ratio will approximately equal the risk ratio. True ***False***
5. In a case-control study, the ratio of cases to controls should always be 1:1. True ***False***

Problem 15

A screening test for a newly discovered disease is being evaluated. In order to determine the effectiveness of the new test, it was administered to 900 workers.

150 of the individuals diagnosed with the disease tested positive. A negative test finding occurred in 60 people who had the disease. A total of 50 persons not diseased tested positive for it.

(a) Draw the appropriate 2×2 table to represent the above information. Double check it to make sure that you have it right.

For each of the following, choose the best answer. You may also want to show your answer in terms of the numbers you entered in the table, for example in the form 60/150.

(b) What is the prevalence of disease cases in the population?

A 77%
B 75%
C 9%
D 23%
e Can't be determined from the information given.

(c) What is the sensitivity of the test?

a 71.4%
B 74.0%
C 92.8%
D 23.3%

(d) What is the specificity of the test?

A 46.9%
B 89.3%
c 92.8%
D 27.8%

(e) What is the overall accuracy of the test?

A 71%
B 75%
C 78%
d 88%

(f) What is the predictive value of a positive test? (Positive predictive value.)

A 53.8%
B 71.1%
C 78.8%
d 75.0%

(g) What is the predictive value of a negative test? (Negative predictive value.)

- [A] 76.7%
- [B] 88.8%
- [c] 91.4%
- [D] 92.8%

(h) Say whether "accuracy" is a good measure of the effectiveness of a diagnostic test. Explain why or why not.

Problem 16

Describe the basic means of surveillance commonly used in public health. Explain how these support or fail to support surveillance of the outbreak of H1N1. [10 minutes]

Problem 17

Write a short answer to each of the following. (You might want to cut-and-paste the list, so that you fill in your answer underneath the corresponding item.)

(a) What is a confounder? What problems or opportunities does confounding produce? Describe some strategies for dealing with the problems or for exploiting the opportunities.

(b) What is the difference between relative and absolute risk. Which is more informative in what situations?

(c) Given an example of a setting in which it would be appropriate to calculate an odds ratio. (You can make up the example, but make it plausible.) Create a table in the standard format that gives an odds ratio of 3. You can make up the numbers any way you want.

(d) Compare and contrast cohort and case-control studies? In what circumstances is a case-control study most appropriate?

(e) What is an "epidemic curve?" Describe how the epidemic curves for a point source and a propagative epidemic differ?

Problem 18

Which study design is best suite for establishing temporality?

- [A] Case-Control
- [B] Cross-Sectional
- [c] Prospective Cohort
- [D] Retrospective Cohort

Problem 19

Cohort studies are prone to bias because of selection. Explain how the healthy-worker effect is a type of selection bias.

Problem 20

In a cohort study, a group of boys aged 8-15 years who were arrested because of substance abuse and identified as chemically dependent, were followed for 15 years. These boys were also classified according to whether they had a history of sexual abuse. Of interest was whether a history of sexual abuse was significantly associated with suicide attempt in these chemically dependent boys, the data for which are presented in the table:

History of Sexual Abuse	Suicide Attempt Yes	No
Yes	11	34
No	29	224

(a) What is the risk of suicide attempt for the chemically dependent boys with a history of sexual abuse?

- [a] $11/45 = 244$ per thousand
- [B] $11/34 = 323$ per thousand
- [C] $29/253 = 115$ per thousand
- [D] $29/224 = 129$ per thousand
- [E] $30/258 = 116$ per thousand

(b) What is the risk of suicide attempt for the chemically dependent boys without a history of sexual abuse?

- [A] $11/45 = 244$ per thousand
- [B] $11/34 = 323$ per thousand
- [c] $29/253 = 115$ per thousand
- [D] $29/224 = 129$ per thousand
- [E] $30/258 = 116$ per thousand

(c) What is the risk ratio?

- [A] 323/129
- [b] 244/115
- [C] 115/129
- [D] $11 \cdot 224/29 \cdot 34$

(d) Suppose eight boys with a history of sexual abuse were lost to follow-up and therefore not reported in the table. In analyzing the situation, you estimate risk ratios for two extreme hypothetical scenarios, one in which all the missing boys attempted suicide, and the other in which none of the missing boys attempted suicide. You get two confidence intervals on the risk ratio:

(i) 1.90 to 5.14

(ii) 0.97 to 3.39.

Which hypothetical scenario does confidence interval (i) correspond to:

- [a] All the missing boys committed suicide.
- [B] None of the missing boys committed suicide.
- [C] No way to know.

Problem 21

A cohort study involving Swedish females born between 1952 and 1989 assessed the association between eating disorders and parental education. From the results of this study presented in the table, calculate an appropriate measure to indicate whether a relationship exists between mother's education and the daughter's risk of having an eating disorder and present your conclusion.

Mother's Education	Number	Percent	Number of Events	Person-Years at Risk
Elementary	2791	21.3	10	42592
Secondary	6365	48.5	21	80468
Post-secondary	3968	30.2	22	43358

We have person-years of exposure for each education level and the number of events, so we can calculate a rate. To compare different education levels, take a rate ratio.

The rates are:

- Elementary: 10/42592, or 2.3 per 10000 person years
- Secondary: 21/80468 or 2.6 per 10000 person years
- Post-secondary: 22/43358 or 5.1 per 10000 person years.

In computing a rate ratio, you need to pick a baseline group. In some sense the choice is arbitrary. Here there is no obvious or intuitive connection between education level (of the parent) and eating disorders, so I'll pick "Elementary" as the base rate.

- Secondary / Elementary: 2.6/2.3 or a rate ratio of 1.13
- Post-secondary / Elementary: 5.1/2.3 or a rate ratio of 2.22 per 10000

These data indicate that daughters of mothers with a post-secondary education have more than twice the relative risk of eating disorders compared to daughters of mothers with an elementary eduction.

Do keep in mind that the absolute risk is small.

Problem 22

A trial is likely to be underpowered, biased, or misleading if which of the following occurs:

A. A large number of participants is lost to follow up.
B. A substantial number of participants do not adhere to the study intervention.
C. Baseline variables are not balanced between intervention groups.
D. All of the above lead to these problems.